Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

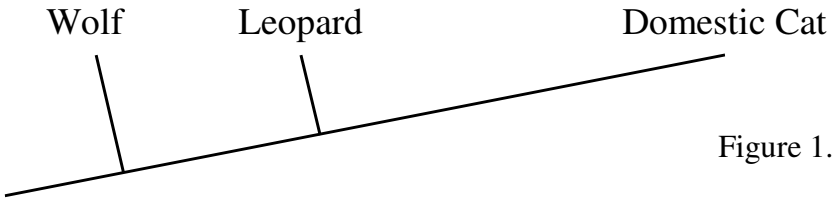

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

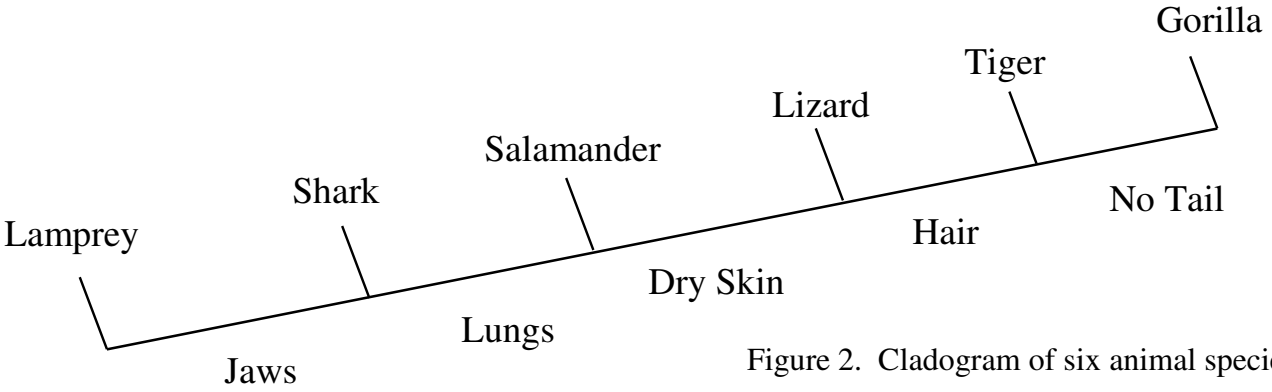

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

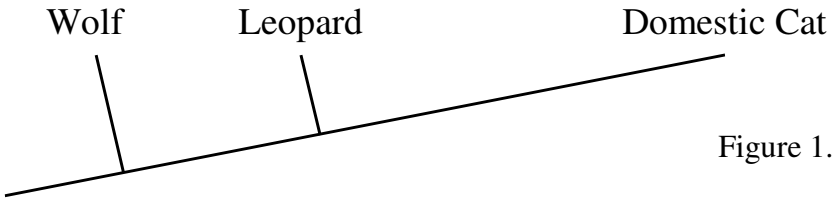

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

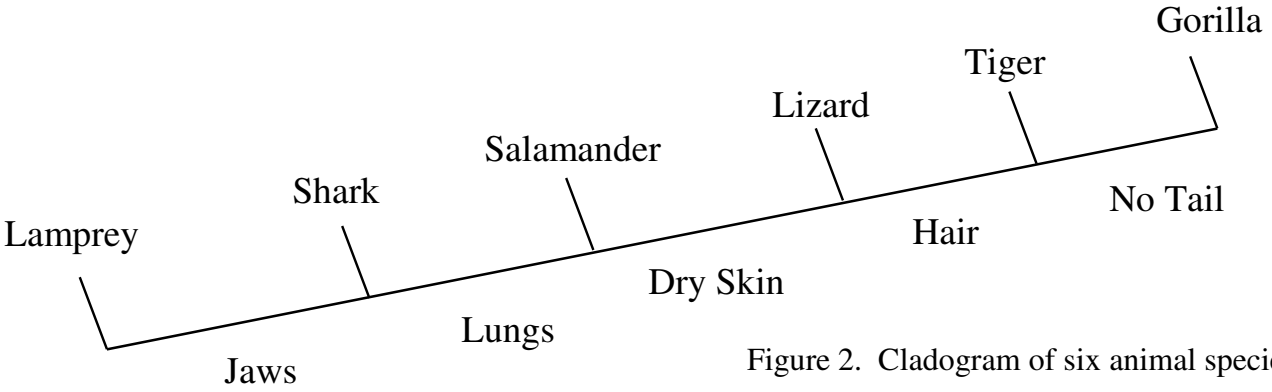

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

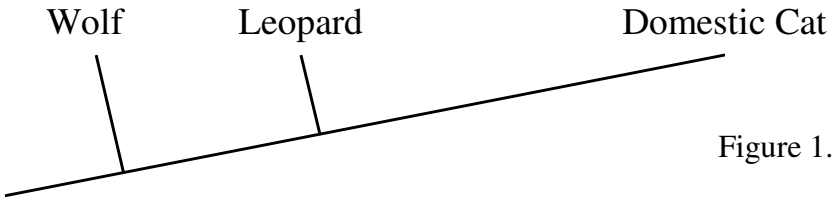

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

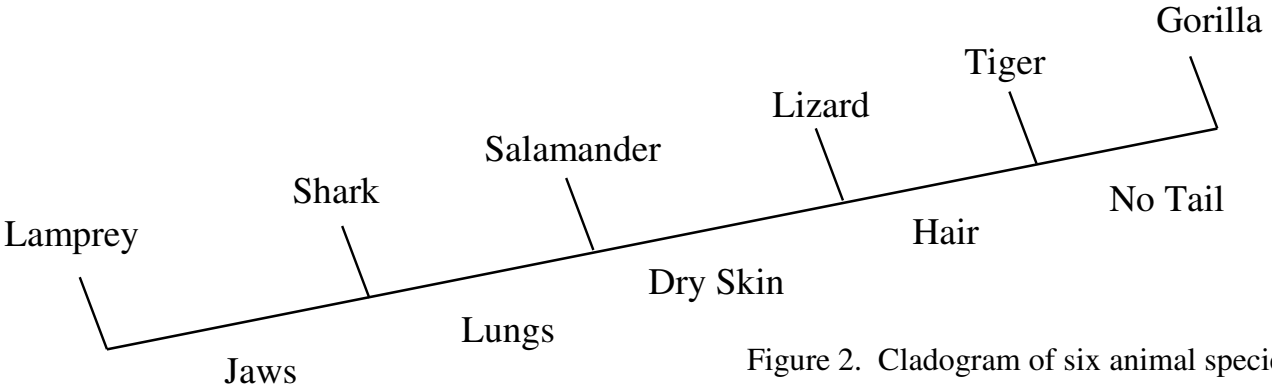

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

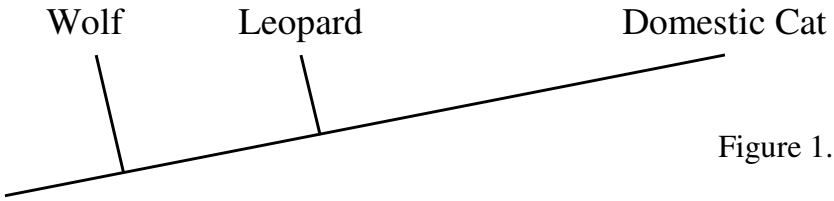

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

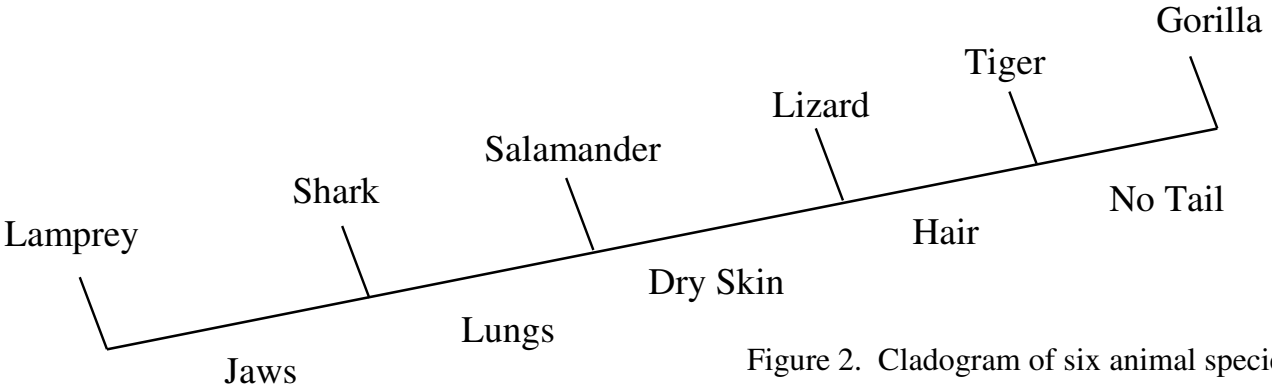

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

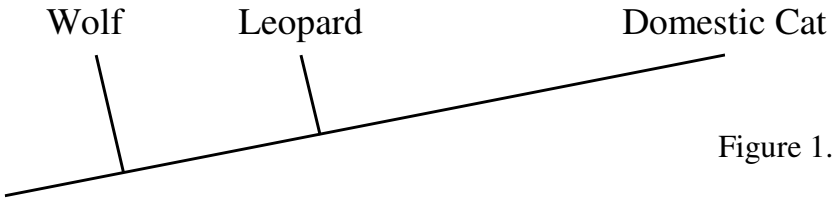

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

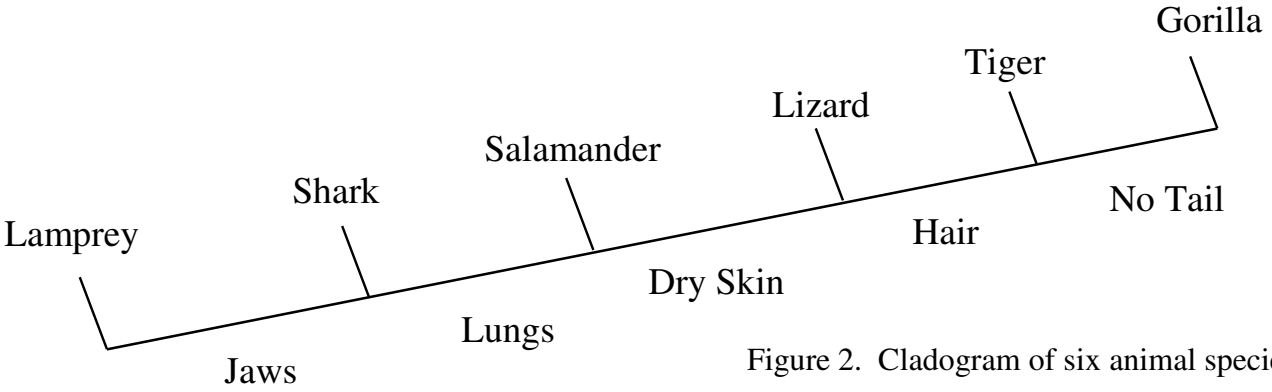

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

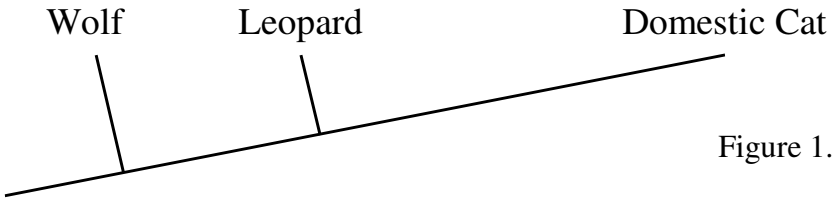

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

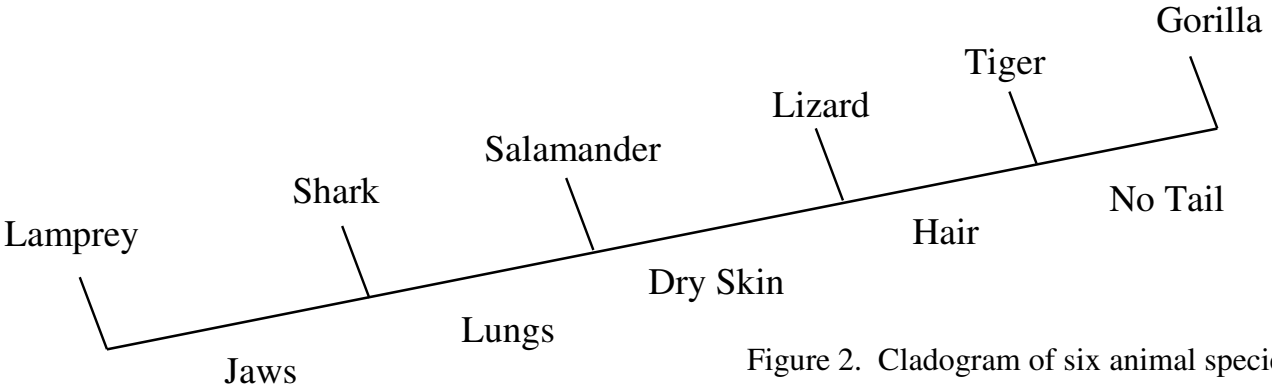

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

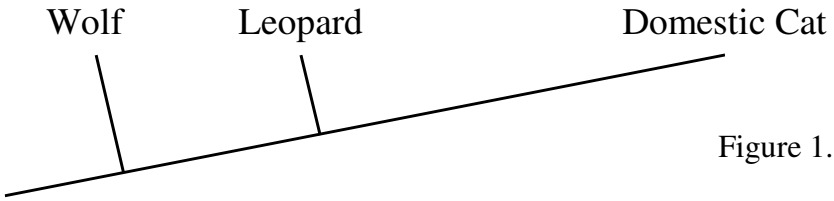

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

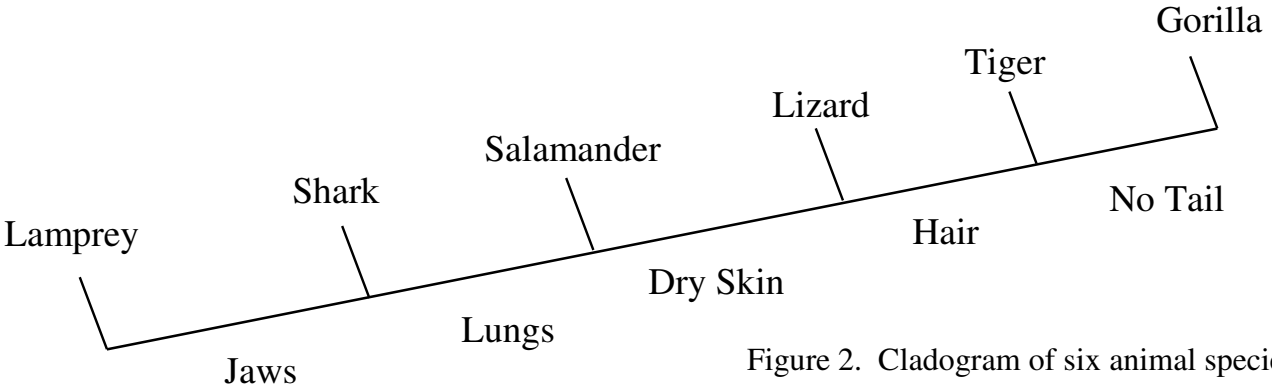

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

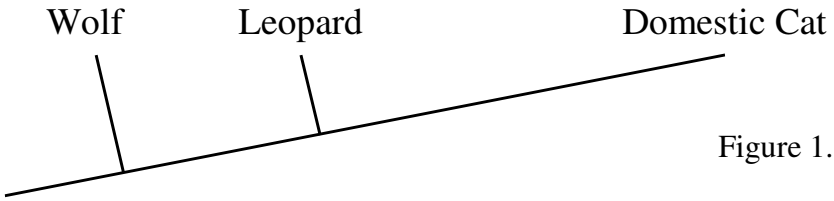

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

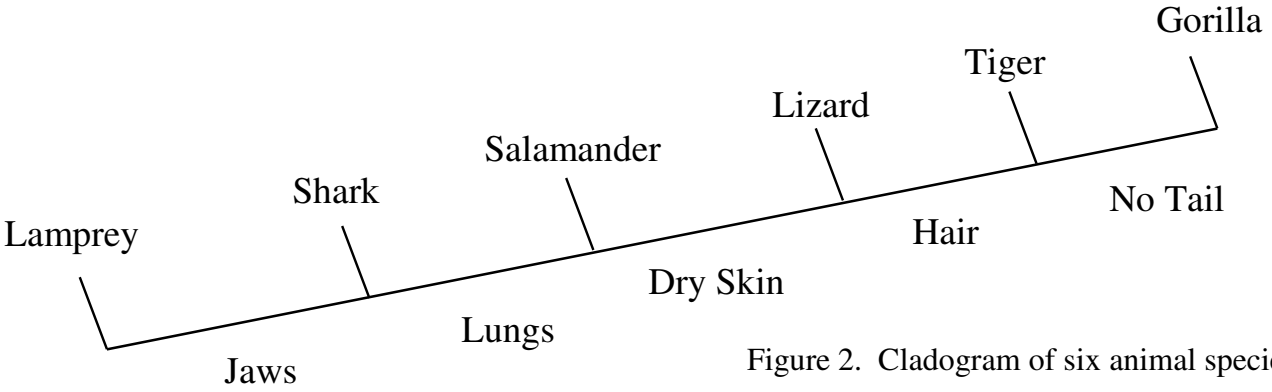

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

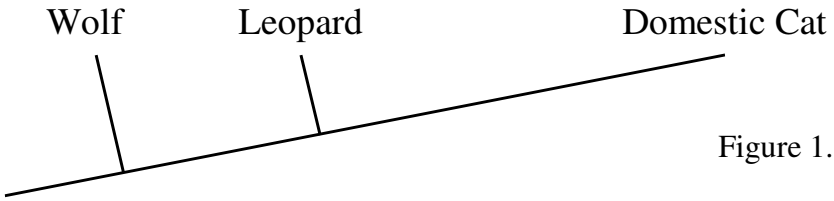

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

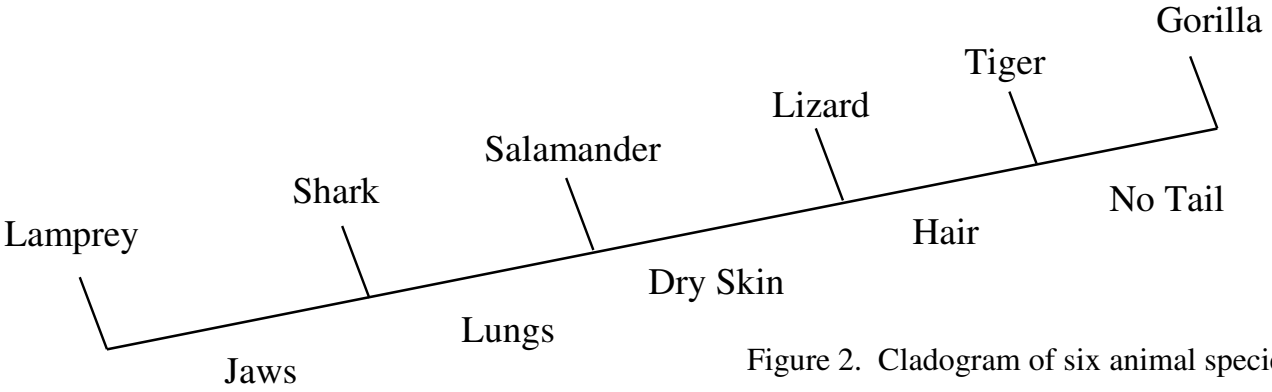

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree

Name _____

Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

Introduction

Bioinformatics is a powerful tool which can be used to determine evolutionary relationships and better understand genetic diseases. You are going to use this tool to explore the conservation of a popular enzyme, cytochrome C, and how it is present in different eukaryotic organisms.

Background

The Human Genome Project (HGP) was completed by scientists in 2003 and was coordinated by the U.S. Department of Energy and National Institutes of Health. The goals of the project were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA.
- Store the genetic sequences in databases.
- Improve tools for data analysis.
- Transfer related technologies to the private sector.
- Address ethical, legal, and social issues arising from the identification of genetic data.

The project mapped not only the genome of humans but also of other species such as *Drosophila melanogaster* (fruit fly), mouse, and *Escherichia coli*. The locations and complete sequences of the genes in each of these species are available for anyone in the world to access on the Internet.

This information is important because the ability to identify the precise location and sequence of human genes will allow greater understanding of genetic diseases. Also, learning about the sequence of genes in other species helps us to understand evolutionary relationships among organisms. Many of our genes are similar if not identical to those found in other species.

For example, a gene in fruit flies is found to be responsible for a particular disease. Scientists might wonder is this gene found in humans and does it cause a similar disease. It would take years to read through the human genome to locate the same sequence of base pairs. Given time constraints, this is not practical—so a technological method was developed.

Bioinformatics is a study that combines statistics, mathematical modeling, and computer science to analyze biological data. Through bioinformatics, entire genomes may be quickly compared in order to detect and analyze their similarities and differences. BLAST (Basic Local Alignment Search Tool) is an extremely useful bioinformatics tool which allows users to input a gene sequence of interest and search entire genomic libraries for identical or similar sequences.

Name _____

Classification of organisms based on evolutionary history is called *phylogenetic systematics*. Scientists study how different organisms are related to determine if they have common ancestry. Today most scientists practice *cladistics*. Cladistics is a taxonomic approach that classifies organisms according to the order in time at which branches arise along a phylogenetic tree without considering the degree of morphological divergence. A phylogenetic diagram based on cladistics is called a *cladogram*. It is a tree constructed from a series of two-way branch points. Each branch point represents the divergence of a common ancestor. The cladogram is treelike where the endpoint of each branch represents a specific species (see Figure 1 below).

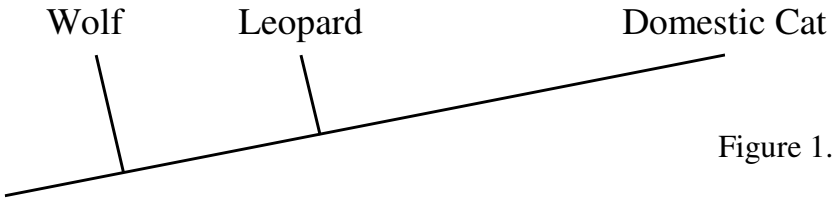

Figure 1. Sample Cladogram

The cladogram featured in Figure 2 includes additional details such as the evolution of particular physical structures called derived characters. Note that the placement of the derived characters corresponds to that character having evolved. Every species **above** the character label possesses that structure. For example, lizards, tigers, and gorillas all have dry skin, whereas salamanders, sharks, and lampreys do not have dry skin.

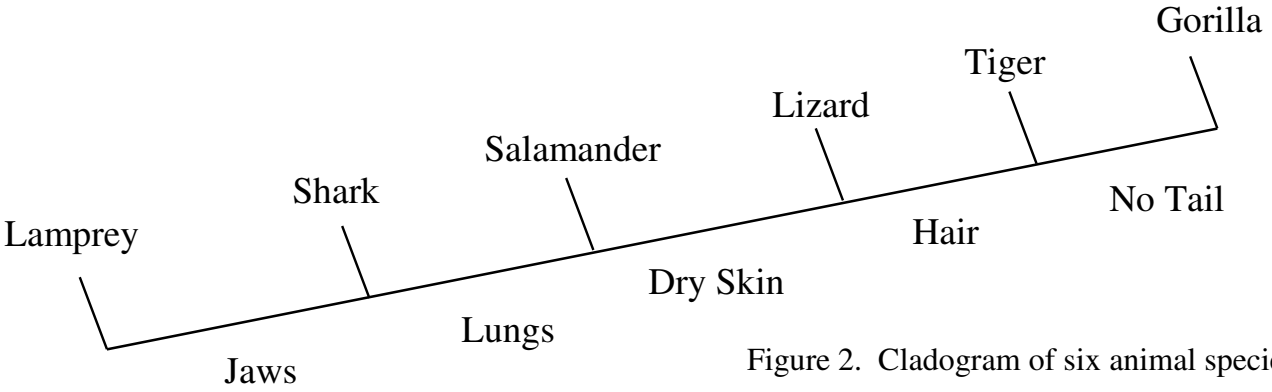

Figure 2. Cladogram of six animal species

Evolutionary changes stemming from random mutations events can alter a protein's primary structure. Some mutations do not allow the organism to survive. In order the change to propagate, the mutation must either allow the organism to have the same evolutionary ability as it had previously or increase its probability to survive and reproduce. Sometimes a mutation can improve the fitness of a host in its natural environment. A classic Darwinian example is sickle cell anemia. This is a result of a single mutation whose adaptive consequences turned out to be beneficial to combat malaria. Normal hemoglobin cells have a high potassium concentration whereas hemoglobin sickle cells do not contain as much potassium. In order for a malaria parasite to survive it needs cells with a high potassium concentration. Thus they do not survive in sickle cells.

Name _____

Pre-Lab Questions

1. Cytochrome c is a highly conserved protein, found in plants, animals, and many unicellular organisms. Do a little research to determine why it is so highly conserved. In your explanation, be sure to include the function of cytochrome C. In order to receive any credit, you must also cite the source where you found this information.
(Note: Use appropriate scientific resources, not websites like Wikipedia or ask.com)

- Source:

2. Draw a cladogram using the information provided in the table below.

		TAXA			
		Pine Trees	Mosses	Flowering Plants	Ferns
Characteristics	Vascular Tissue	1	0	1	1
	Flowers	0	0	1	0
	Seeds	1	0	1	0

Table 1. Character Table. A zero (0) indicates that a character is absent; a one (1) indicates that a character is present.

Name _____

Procedure

Part A. Gathering cytochrome C sequences from NCBI

1. Log in to your school gmail and create a Google Drive document in order to save the sequences you will be gathering.
    - Save the document as "yourlastname.cytochrome C sequences"
2. Go to the National Center for Biotechnology Information Website: http://ncbi.nlm.nih.gov
3. Change the "All Databases" drop down menu to "Protein"
4. In the search box type in "cytochrome C" and click the search button
    - You will then see a list of information for this protein in various organisms. You will narrow your search by identifying specific organisms according to the list found below
5. In the search box type "cytochrome C horse." Click the search button.
6. Many choices will appear. Use the first record that appears.
    - For the horse (*Equus caballus*), the first record has the following accession number: "NP_001157486.1"
7. Click the link under this record that says "FASTA." This will list the amino acid sequence in a simple format.
8. Copy and paste the amino acid sequence into your Google Drive document
    - Be sure to include the species name and accession number used
    - You MUST also include the ">" symbol when you copy and paste
    - You can then type in the common name (horse) between this symbol and the accession number in your Goggle Drive document
9. Go back to the NCBI Website and arrow back until you arrive at the protein search page again. Backspace to remove the organism "horse" and then type the next organism from the list below
    - Note: Cytochrome C should still precede the organism in the search box.
10. Continue copying and pasting the cytochrome C sequences until you have gathered them for all 15 of the following organisms

- Horse
- Chicken
- Zebrafish
- Human
- Chimpanzee
- Rabbit
- Rattlesnake
- Bullfrog
- Dog
- Bee
- Cow
- Baboon
- Mouse
- Fruitfly
- Plant

11. Be sure to label the name and accession number for each organism in your Google Drive document
12. Print a copy of your document (and be sure to turn that page in with your completed lab) and also share your document with me (pamos@fortcherry.org)
13. Manually compare the amino acid sequences of the following 4 organisms to that of the **human** and count the number of differences between the **human** cytochrome C sequence and the four others and record in the table in the results section

- Horse
- Baboon
- Cow
- Chimpanzee

Name _____

Part B. Comparative Genomics and Bioinformatics

1. We are going to use Clustal Omega to help compare the sequences you found
    - Go to http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Make sure the drop down menu under Step 1 has "PROTEIN" selected
3. Copy your entire list of amino acid sequences for all 15 of your organisms from your Google Drive Document (with ">" and labels included!) and paste them into the text box found in Step 1.
4. The Output Format under Step 2 should have "Clustal w/o numbers" selected
5. Click Submit
6. After a few moments, your CLUSTAL multiple sequence alignment should appear.
7. You can click on "Show Colors" for a colorful comparison. The symbols below the sequences also indicate similarities.
8. Click on "Phylogenetic Tree"
9. Draw this phylogenetic tree in the space provided in your results section below (you do NOT need to include the numbers)
10. Answer the analysis questions based on your results
11. Complete the Extension activity using the gene assigned to you

Results

Organism	Number of Different Amino Acids compared to cytochrome C sequence of Humans
Horse	
Baboon	
Cow	
Chimpanzee	

Phylogenetic Tree

Name _____

Analysis

Compare the results from your table (where you manually counted the number of differences in the amino acid sequences) with the phylogenetic tree that was produced using the BLAST program. (Note: For the following questions, you are ONLY focusing on the **horse**, **baboon**, **cow**, and **chimpanzee** in comparison to humans and each other)

1. Which of the 4 organisms has the most similarities in its cytochrome c amino acid sequence compared to humans?

2. Which organism has the next most similar cytochrome c amino acid sequence compared to humans?

3. How are these similarities reflected in the resulting phylogenetic trees?

4. Which 2 organisms have the most differences in their cytochrome c amino acid sequence compared to humans?

5. How is this reflected in the resulting phylogenetic trees?

Name _____

Extension: Comparing Other Genes

What other genes are conserved among different species and what does this suggest about their evolutionary relationships? You will investigate these questions by researching the function of the protein created from an assigned gene and comparing its amino acid sequence in 10 total organisms.

- Record your assigned gene in the space provided and research the function of the resulting protein, making sure to cite your source.
- **In addition to humans**, select 9 more of the 15 organisms used when analyzing the cytochrome C sequences and record their names below.
- Repeat the procedure used in Parts A and B for those **10 total organisms**
    - Create a new Google Drive document titled "yourlastname.yourassignedgene sequences"
        - Save, print, and share the sequences for humans and the 9 other organisms you selected as well as pasting them into the Clustal Omega site to produce phylogenetic trees
    - Draw the resulting phylogenetic tree on the next page

Assigned Gene: _____

Function (be sure to cite the source):

Name _____

Organisms selected for comparison:
1. Humans
2.
3.
4.
5.
6.
7.
8.
9.
10.

Phylogenetic Tree